

Sequential Use of Neural Networks for Survival Prediction in AIDS

Lucila Ohno-Machado, M.D., Ph.D.

Decision Systems Group, Brigham and Women's Hospital, Harvard Medical School
75 Francis Street, Boston, MA 02115 — machado@dsg.harvard.edu

Prognostic assessment of patients is a key part of medical care. Although neural networks can be used to model survival, their accuracy has been limited for a variety of factors, including (1) the lack of data balance in certain intervals and (2) the lack of representation of temporal dependencies in the network architecture. Both problems can be solved with the use of sequential neural networks, which establish predictions for a certain time point and then use these predictions to produce survival estimates for other time points. If the sequence of models is adequate, sequential neural networks produce more accurate estimates of survival than standard neural networks, as shown in this example in the domain of AIDS. Assessments of survival in one, two, three, five and six years become more accurate (as measured by the areas under the ROC curves) when initial predictions of survival in four years are used in a sequential neural network model.

OUTCOME PREDICTION OVER TIME

Prognostic information can help (1) medical researchers to identify certain patterns of disease progression and design experiments, (2) health maintenance organizations and other members of the health care industry to predict the needs of their served populations and to allocate resources accordingly, and (3) patients to understand and make informed decisions about their conditions.¹

The increasing availability of electronic medical data calls for statistical and machine-learning methods that are able to extract important information from patient records and develop accurate predictive models. The most widespread application of statistical classification for medical prognosis has been in predicting outcomes and establishing severity indices for patients in intensive care units (ICUs). The APACHE system² uses logistic regression to determine factors that are correlated with good prognosis and to make outcome predictions for patients in ICUs. Severity indices for prognosis of trauma, such as TRISS and ASCOT, also use logistic regression models.³ The majority of currently available prognostic systems were developed for acute conditions. Machine-learning methods have not been widely used to make prognosis in medicine. A few neural network applications are exceptions.^{3,4,5,6}

Neural networks for prognosis

Neural networks are computational models that may

be used in the same tasks as regression models. They have been used extensively for medical diagnosis in the past decade,⁷ but their use for medical prognosis has been more limited, as opposed to classic statistical methods. From a set of existing data, a model is constructed and the parameters are estimated. The back-propagation algorithm⁸ has been used extensively to estimate parameters in a neural network. These parameters are called *weights*. Recent studies have compared neural networks to several statistical classification methods⁹ and outlined the similarities in scope and purpose of neural networks and logistic regression models.¹⁰

Studies that use neural networks to establish prognosis for groups of patients usually provide outcome predictions for a single point in time. Utilizing aggregates of single point estimates to predict individual survival and to delineate patterns of disease progression may yield spurious results, including nonmonotonic survival curves, as indicated in Figure 1. These results occur because predictions tend to be less accurate in intervals where data are unbalanced (i.e., the ratio between deaths and all cases is close to 0 or 1),¹¹ and no provisions for accounting for internal time dependencies are made in isolated models.

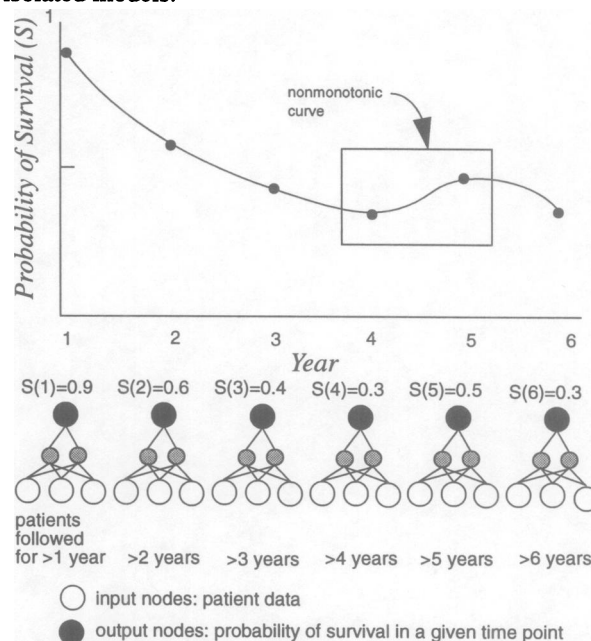


Figure 1. Standard neural networks for prognosis. Combination of isolated predictions from standard neural networks may lead to spurious survival curves.

Sequential Neural Networks

Sequential neural network systems are constructed incrementally. In each step of the sequence, predictions for one time point are produced by a neural network. These predictions are passed forward to other networks in the system. One or more networks may provide predictions that become inputs for other networks in the sequence. The result is a chain or a hierarchy of neural networks, as shown in a simplified form in Figure 2.

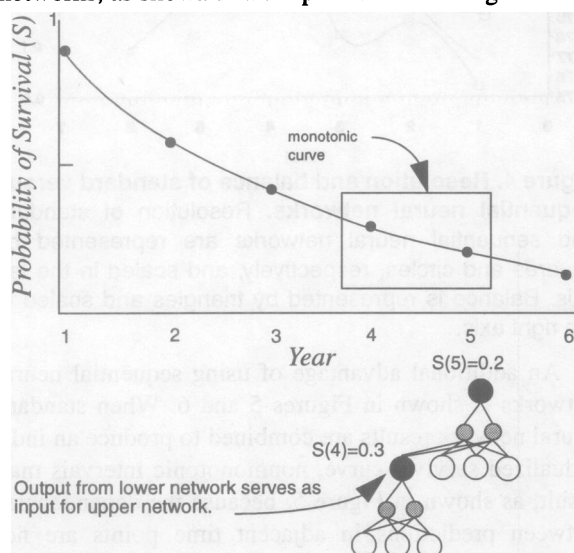


Figure 2. Sequential neural networks for prognosis. In sequential neural networks, dependencies among time point predictions are explicitly represented, resulting in monotonic survival curves. In this example, predictions for year 4 are entered in the model that predicts survival in year 5.

A sequential system of neural networks has been shown to provide predictions that are more accurate in terms of calibration and resolution than the ones produced by standard neural networks in the domain of coronary heart disease.¹² In this work, sequential neural networks are used to model survival with AIDS.

PREDICTION OF SURVIVAL WITH AIDS

Neural networks have rarely been used to model survival in the domain of HIV infection. The exceptions relied solely on standard neural networks. The goal of this study was to produce a model of disease progression in AIDS using sequential neural networks, and compare the model's accuracy with that of a model constructed using only standard neural networks.

A subset of the ATHOS data¹³ set was used for all experiments. Not all AIDS patients from the ATHOS data set were used because some lacked the date of AIDS diagnosis. Table 1 shows the distribution of cases according to the years of follow-up. Censored cases were not used in the models, so the total number of

patients available for each year varies. The balance of cases in each set is defined as

$$\text{balance} = \min(\text{Deaths}/\text{Total}, \text{Survivals}/\text{Total})$$

A balance of 0.5 means that the proportion of cases is 1:1, and is the one that best facilitates classification by statistical and neural network models.

Table 1. Distribution of cases according to year of follow-up.

Year of follow-up	Dead	Alive	Total	Balance
1	64	850	914	0.0752
2	150	606	756	0.1984
3	229	358	587	0.3901
4	257	199	456	0.4364
5	274	86	360	0.2388
6	277	28	305	0.0918

The major endpoint in this analysis was prediction of mortality due to AIDS-related conditions, measured from the date of AIDS diagnosis using the 1993 CDC definition.¹⁴ Variables were included in the model only when the literature showed that they have been proven to be informative. Not all published markers for disease progression in HIV infection were available in the ATHOS data set. Only baseline values, at the time of AIDS diagnosis, were used.

Demographic and socioeconomic explanatory variables included age, gender, race, risk group, AIDS-defining diagnoses, insurance coverage, length of stay in hospital, and time elapsed from the estimated HIV seroconversion. Clinical findings included fatigue, weight loss, diarrhea, mental status, and Karnofsky scores. Laboratory test results included CD4 counts, CD4/CD8 ratio, hemoglobin, erythrocyte sedimentation rate, erythrocyte and platelet counts, white blood cell counts, serum p24 antigen, serum β -2 microglobulin, total cholesterol, HDL, and albumin levels. Antiretroviral and prophylactic medications for opportunistic infections and AIDS-related conditions reported after the patient entered the study were also recorded. Continuous variables were represented as such, but they were normalized before entry. Dummy coding was used for categorical variables.

Standard and sequential neural networks were constructed to model survival in six time points (year one, year two, and so forth). Each network had a single output representing survival at a certain time point. The standard neural networks had 38 inputs and 20 hidden nodes. The sequential neural networks had 39 inputs (one extra input representing predictions in another time point). All networks were trained by backpropagation with adaptive learning rate. Overfitting was monitored in a holdout set of 40 percent of the cases. The software

package NevProp2¹⁵ was used. Cases were divided into training and test sets using the bootstrap method.¹⁶ Accuracy was measured in terms of resolution, or discriminatory ability, of the neural network models. Resolution was defined in terms of the area under the receiver operating characteristic (ROC) curve.¹⁷

RESULTS

A comparison of resolution obtained by standard and sequential neural networks in predicting survival with AIDS at one year from the diagnosis is shown in Figure 3. By using the predictions of year four as inputs, the sequential neural network could discriminate patients who die and patients who survive more accurately than the standard network. Similar improvements occurred for other time points, as shown in Table 2.

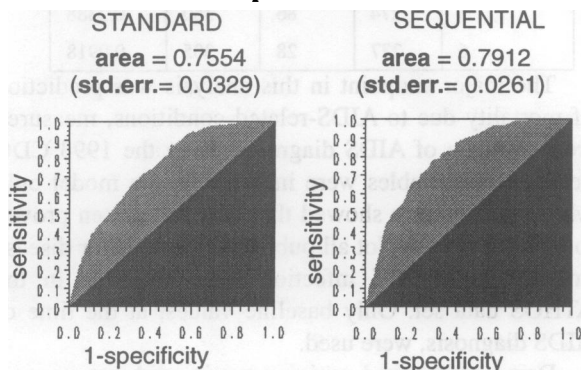


Figure 3. Change in resolution using standard and sequential neural networks at one year. Resolution of predictions of survival in one year resulting from the use of sequential neural networks was significantly higher than that of standard neural networks.

Table 2. Resolution of standard and sequential neural network models.

Year	Standard		Sequential	
	Area ROC	st. error	Area ROC	st. error
1	0.7554	0.0329	0.7912	0.0304
2	0.7879	0.0213	0.7936	0.0221
3	0.7818	0.0197	0.8063	0.0198
4	0.8703	0.0174	0.8703	0.0174
5	0.8647	0.0207	0.8720	0.0198
6	0.8346	0.0317	0.9065	0.0289

As illustrated in Figure 4, resolution correlated positively with data balance in standard models: the more balance between survivals and deaths in a given time point, the higher the resolution. Although sequential neural network resolution also depended on the balance of data, resolution was higher than that of standard mod-

els for unbalanced time points (e.g., one and six years) for $\alpha=0.10$.

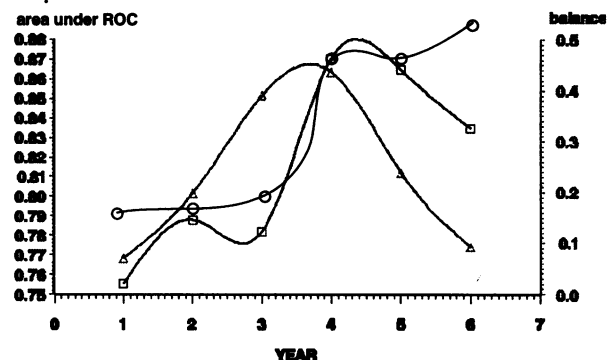


Figure 4. Resolution and balance of standard versus sequential neural networks. Resolution of standard and sequential neural networks are represented by squares and circles, respectively, and scaled in the left axis. Balance is represented by triangles and scaled in the right axis.

An additional advantage of using sequential neural networks is shown in Figures 5 and 6. When standard neural network results are combined to produce an individualized survival curve, nonmonotonic intervals may result, as shown in Figure 5, because the dependencies between predictions in adjacent time points are not taken into account.

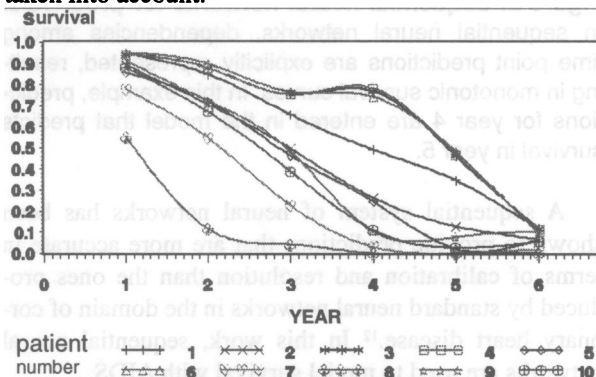


Figure 5. Combination of standard neural network results. Since the models do not relate to each other, survival curves that are not monotonically decreasing, although impossible in theory, can be produced, such as for patient number 4.

If sequential neural networks are used, the number of nonmonotonic intervals tends to decrease, because certain time point dependencies are accounted for. Figure 6 shows an example of 10 survival curves produced by sequential neural network models that have predictions of year 4 as inputs.

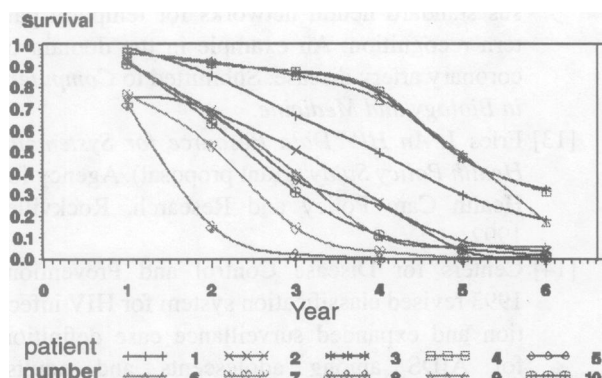


Figure 6. Sequential neural networks. Survival curves produced by the sequential models tended to have fewer nonmonotonic intervals.

DISCUSSION

Prediction of outcomes for an individual patient depends on several variables. Unknown interactions, as well as noise, may influence the results. Although neural networks have been shown to be resilient to noise and able to handle interactions,¹⁸ their predictive accuracy is severely limited when the data are not well balanced (i.e., the priors for some outcome classes are low). This limitation is not exclusive to neural networks, and current methods for decreasing its impact on classification accuracy have been applied to other classification systems as well: equalization of priors (by sampling the training set in a way that would make representation of classes more balanced) or application of cost functions (or utility functions) in parameter estimation.¹⁹ The problem with the first approach is that information about prior probabilities is lost; the problem with the second approach is that every time the functions change, new models have to be created.²⁰

The assessment of prognosis for patients over time illustrates the need for dealing with the problem of unbalanced data. At the extremes of time intervals that represent duration of disease or life span of a human being, there are time points in which the data represent few people with or without a certain condition (e.g., dead, in the case of initial time points in a study of survival). In these cases, the classification of infrequent exemplars is difficult. Sequential application of neural networks to partial subtasks facilitates recognition of these infrequent cases and promotes an increase in total classification accuracy. Sequential application of neural networks to the prognosis of patients who have AIDS provides results that not only are more accurate in terms of discrimination (especially for infrequent cases) but are also more realistic because they incorporate the commonsense knowledge that predictions of survival are necessarily correlated over time. Sequential methods

make more use of the available information and can significantly enhance the predictive ability of current prognostic models, delineating patterns of disease progression that could not be envisioned by current methods. This increase in predictive ability has the potential to (1) empower patients, since they will have more precise estimates as to how their disease will progress; (2) empower healthcare givers, who will be able to make more informed decisions on the course of therapeutics; and (3) empower health care organizations, which will be able to anticipate the needs of their covered population and anticipate costs.

Sequential neural networks are easy to build, but certain sequential models may require longer training times than their conventional counterparts. Only two-step sequential systems are described in this study. The use of more steps implies using more computer resources, but has the potential to improve further the resolution of certain models. For example, a three-step model that first predicts the extremes of an interval and then applies those predictions to the model that predicts the middle of the interval may improve resolution.

CONCLUSION

Survival analysis can be viewed as a problem in which rare categories of events need to be discriminated. Standard neural networks can be accurate predictors when the frequency of events is not low. Sequential neural networks provide a way to achieve high accuracy even for low-frequency events.

A sequential system of neural networks presented here offers a solution to the problem of recognizing infrequent patterns in survival data. The sequential system makes use of accurate predictions for a certain time point to develop a model that makes predictions for other time points in which the accuracy is not as high. The use of this type of information not only allows an increase in resolution for certain time points, but also increases the model's overall consistency, by producing survival curves that have fewer nonmonotonic intervals.

Acknowledgments

This work was developed at the Section on Medical Informatics, Stanford University School of Medicine as part of my Ph.D. thesis. I thank Mark Musen, Ted Shortliffe, David Rumelhart, Michael Walker, and Alan Garber for useful discussions in different aspects of the present work. This work has been funded by the Conselho Nacional de Pesquisa (CNPq), Brazilian Ministry of Education. Computing facilities were provided by CAMIS through grant LM05305 from the National Library of Medicine.

Reference

- [1] Justice AC; Feinstein AR; Wells CK. A new prognostic staging system for the acquired immunodeficiency syndrome. *New England Journal of Medicine*, 1989 May 25, 320(21):1388-93.
- [2] Knaus WA; Wagner DP; Draper EA; Zimmerman JE; Bergner M; Bastos PG; Sirio CA; Murphy DJ; Lotring T; Damiano A. The APACHE III prognostic system. Risk prediction of hospital mortality for critically ill hospitalized adults. *Chest*, 1991 Dec, 100(6):1619-36.
- [3] McGonigal MD; Cole J; Schwab CW; Kauder DR; Rotondo MF; Angood PB. A new approach to probability of survival scoring for trauma quality assurance. *Journal of Trauma*, 1993 Jun, 34(6):863-8.
- [4] Ravdin PM; Clark GM. A practical application of neural network analysis for predicting outcome of individual breast cancer patients. *Breast Cancer Research and Treatment*, 1992, 22(3):285-93.
- [5] Tu JV; Guerriere MR. Use of a neural network as a predictive instrument for length of stay in the intensive care unit following cardiac surgery. *Computers and Biomedical Research*, 1993 Jun, 26(3):220-9.
- [6] Ohno-Machado L; Musen MA. A comparison of two computer-based prognostic systems for AIDS. *Proceedings of the Nineteenth Symposium on Computer Applications in Medical Care*, Hanley and Belfus, Philadelphia, 1995.
- [7] Reggia JA. Neural computation in medicine. *Artificial Intelligence in Medicine*, 1993, 5(2):143-57.
- [8] Rumelhart DE; Hinton GE; Williams RJ. Learning internal representation by error propagation. In Rumelhart, D.E., and McClelland, J.L. (eds) *Parallel Distributed Processing*. MIT Press, Cambridge, 1986.
- [9] King RD; Feng C; Sutherland A. Statlog: Comparison of classification algorithms on large real-world problems. *Applied Artificial Intelligence*, 1995, 9:289-333.
- [10] Cheng B; Titterton DM. Neural networks: A review from a statistical perspective. *Statistical Science*, 1994, 9(1):2-54.
- [11] Ohno-Machado L. Identification of low frequency patterns in backpropagation neural networks. In Ozbolt JG(ed), *Proceedings of the Eighteenth Symposium on Computer Applications in Medical Care*, Hanley and Belfus, Philadelphia, 1994.
- [12] Ohno-Machado L; Musen MA. Sequential versus standard neural networks for temporal pattern recognition: An example in the domain of coronary artery disease. Submitted to *Computers in Biology and Medicine*.
- [13] Fries J. *An HIV Data Resource for Systematic Health Policy Study* (grant proposal). Agency for Health Care Policy and Research, Rockville, 1992.
- [14] Centers for Disease Control and Prevention. 1993 revised classification system for HIV infection and expanded surveillance case definition for AIDS among adolescents and adults. *MMWR*, 1992 Dec 18, 41(RR-17):1-19.
- [15] Goodman P. Personal communication, 1994.
- [16] Efron B. Bootstrap methods: Another look at the jackknife. *Annals of Statistics*, 1979, 7:1-26.
- [17] Hanley JA; McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, 1982 Apr, 143(1):29-36.
- [18] Hertz JA; Palmer RG; Krogh, AS. *Introduction to the Theory of Neural Computation*. Addison-Wesley, Redwood City, 1991.
- [19] Lowe D; Webb AR. Exploiting prior knowledge in network optimization: An illustration from medical prognosis. *Network*, 1990, 1:299-323.
- [20] Ohno-Machado L. *Medical Applications of Artificial Neural Networks: Connectionist Models of Survival*. Ph.D. thesis. Departments of Computer Science and Medicine. Stanford University, 1996.